



De-ID
DATA
CORP

De-ID™ Software and caBIG

Dan Wasserstrom

President, De-ID Data Corp

Steven Merahn, MD

Chief Medical Officer

De-ID

DATA

CORP

Features of De-ID

- Accurate, reliable automated de-identification
- Very simple to install and use
 - Network-level operability; batch processing
- Works with word processing files, tab- or comma-delimited or XML formatted data
- Meets all HIPAA ‘safe harbor’ guidelines; option of limited data sets and custom fields

De-ID

DATA

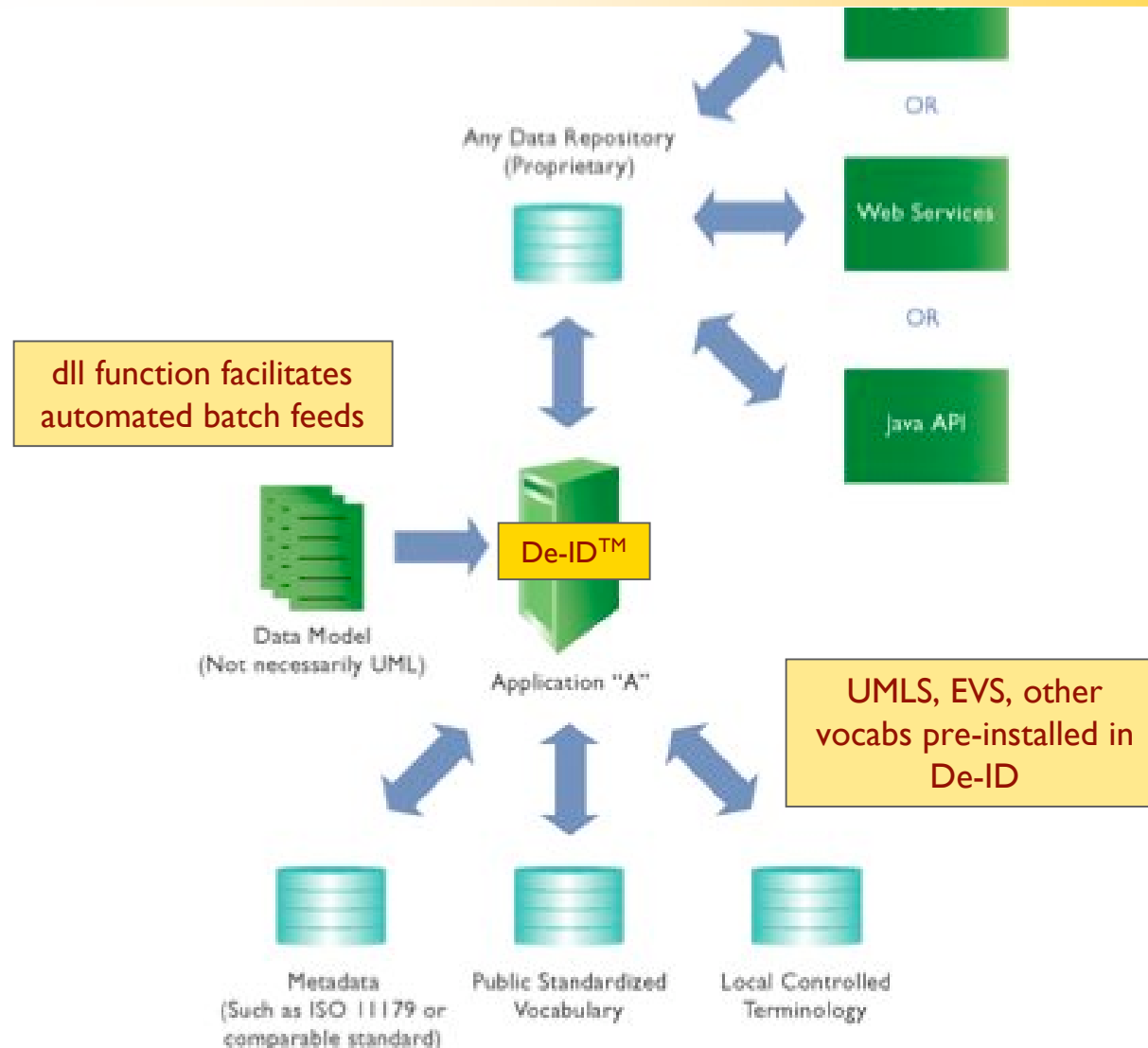
CORP

Why Use De-ID™?

Interface Integration	De-ID can be pre-set and ‘called’ from any application for automated batch processing; can work ‘behind the scenes’
Vocabularies/ Terminologies	UMLS, EVS pre-loaded; institution or geography specific glossaries easily loaded
Data elements	Files must be de-identified before leaving ‘home’ for processing
Information Models	De-identification required before topic mapping or other forms of information modeling

De-ID
DATA
CORP

Compatibility Compliance



De-ID

DATA

CORP

Compatibility Compliance

- Pre-loaded with UMLS, EVS
 - Can also incorporate specialized acronyms and institution-specific terminology; identifiers can be managed through customized rule sets
- Automated .dll 'call' from any application for batch file loading
 - Admin-level settings; C/C++ header files are created in the subfolder of the install directory.
 - Creates deidentification threads that can be viewed 'in progress' or operate behind the scenes
- Output as independent XML; other file formats as needed
- Outputs automatically fed to content repository or 'next step' application (tagging, parsing)

De-ID

DATA

CORP

Benefits of De-ID

- To increase speed and lower costs of “data deposits”
- To maintain consistency of de-identification schema
- To enhance the data value of de-identified records
- Enhances power and usability of traditional de-identification via proxies and offsets for safe-harbor elements, which can be used as data points



De-ID
DATA
CORP

De-ID: A Look Under the Hood

The Manual View

De-ID

DATA

CC



De-ID / Software for de-identifying narrative medical documents



File Options Help

De-identification/output options Ctrl+D

90:
mas Jones
year-old ge
ber 23, 200
sity of Pitts

File Type: N/A

771D-NO
NAME[AAA
CE[pe 2004
DATE[May
NSTITUTE

☒ Output file (de-identified)



☒ Linkage file

Passphrase:



☒ De-identification log

[Safe-harbor compliant Text \(multi-line header\)](#)



Begin de-identification

Available Header Fields:

Header Field	Type
--------------	------

Selected Fields:

Header Field	Type
--------------	------

De-ID copyright (c) 1999-2004, University of Pittsburgh. All rights reserved.

Ver 6.00 (Jan 11 2005)

CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA
CORP

De-ID options: The 18 HIPAA identifiers [CFR 164.514(b)(2)(i) and 164.514(c)(2)]

<input checked="" type="checkbox"/> Names [1] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Patient, relative, employer or household members [1]<input checked="" type="checkbox"/> Health care providers [1]	<input type="checkbox"/> License numbers [11] and vehicle identifiers [12]
<input checked="" type="checkbox"/> Geographical subdivisions smaller than a state [2] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Street addresses [2]<input checked="" type="checkbox"/> Town names and zip codes [2]<input checked="" type="checkbox"/> Hospital names [2]	<input checked="" type="checkbox"/> Device identifiers and serial numbers [13]
<input checked="" type="checkbox"/> Dates and ages <ul style="list-style-type: none"><input checked="" type="checkbox"/> All elements of date, except year [3]<input checked="" type="checkbox"/> Ages under 90 years [not req'd]<input checked="" type="checkbox"/> Ages 90 years and over [3]	<input checked="" type="checkbox"/> Web universal resource locators (URLs) [14] and internet protocol (IP) addresses [15]
<input checked="" type="checkbox"/> Telephone [4] and fax [5] numbers	<input checked="" type="checkbox"/> Biometric identifiers [16] and full-face photographs [17]
<input checked="" type="checkbox"/> Electronic mail addresses [6]	<input checked="" type="checkbox"/> Any other uniquely identifying number, characteristic or code [18] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Pathology specimen numbers [18?]<input type="checkbox"/> Occupations (not of health care provider) [18?]
<input checked="" type="checkbox"/> Social security [7], medical record [8], health plan [9] and other [10] account numbers	

Text (multi-line header)

Safe-harbor defaults Limited dataset defaults

OK

CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA
CORP

De-ID options: The 18 HIPAA identifiers [CFR 164.514(b)(2)(i) and 164.514(c)(2)]

<input checked="" type="checkbox"/> Names [1] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Patient, relative, employer or household members [1]<input checked="" type="checkbox"/> Health care providers [1]	<input type="checkbox"/> License numbers [11] and vehicle identifiers [12]
<input checked="" type="checkbox"/> Geographical subdivisions smaller than a state [2] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Street addresses [2]<input checked="" type="checkbox"/> Town names and zip codes [2]<input checked="" type="checkbox"/> Hospital names [2]	<input checked="" type="checkbox"/> Device identifiers and serial numbers [13]
<input checked="" type="checkbox"/> Dates and ages <ul style="list-style-type: none"><input checked="" type="checkbox"/> All elements of date, except year [3]<input checked="" type="checkbox"/> Ages under 90 years [not req'd]<input checked="" type="checkbox"/> Ages 90 years and over [3]	<input checked="" type="checkbox"/> Web universal resource locators (URLs) [14] and internet protocol (IP) addresses [15]
<input checked="" type="checkbox"/> Telephone [4] and fax [5] numbers	<input checked="" type="checkbox"/> Biometric identifiers [16] and full-face photographs [17]
<input checked="" type="checkbox"/> Electronic mail addresses [6]	<input checked="" type="checkbox"/> Any other uniquely identifying number, characteristic or code [18] <ul style="list-style-type: none"><input checked="" type="checkbox"/> Pathology specimen numbers [18?]<input type="checkbox"/> Occupations (not of health care provider) [18?]
<input checked="" type="checkbox"/> Social security [7], medical record [8], health plan [9] and other [10] account numbers	

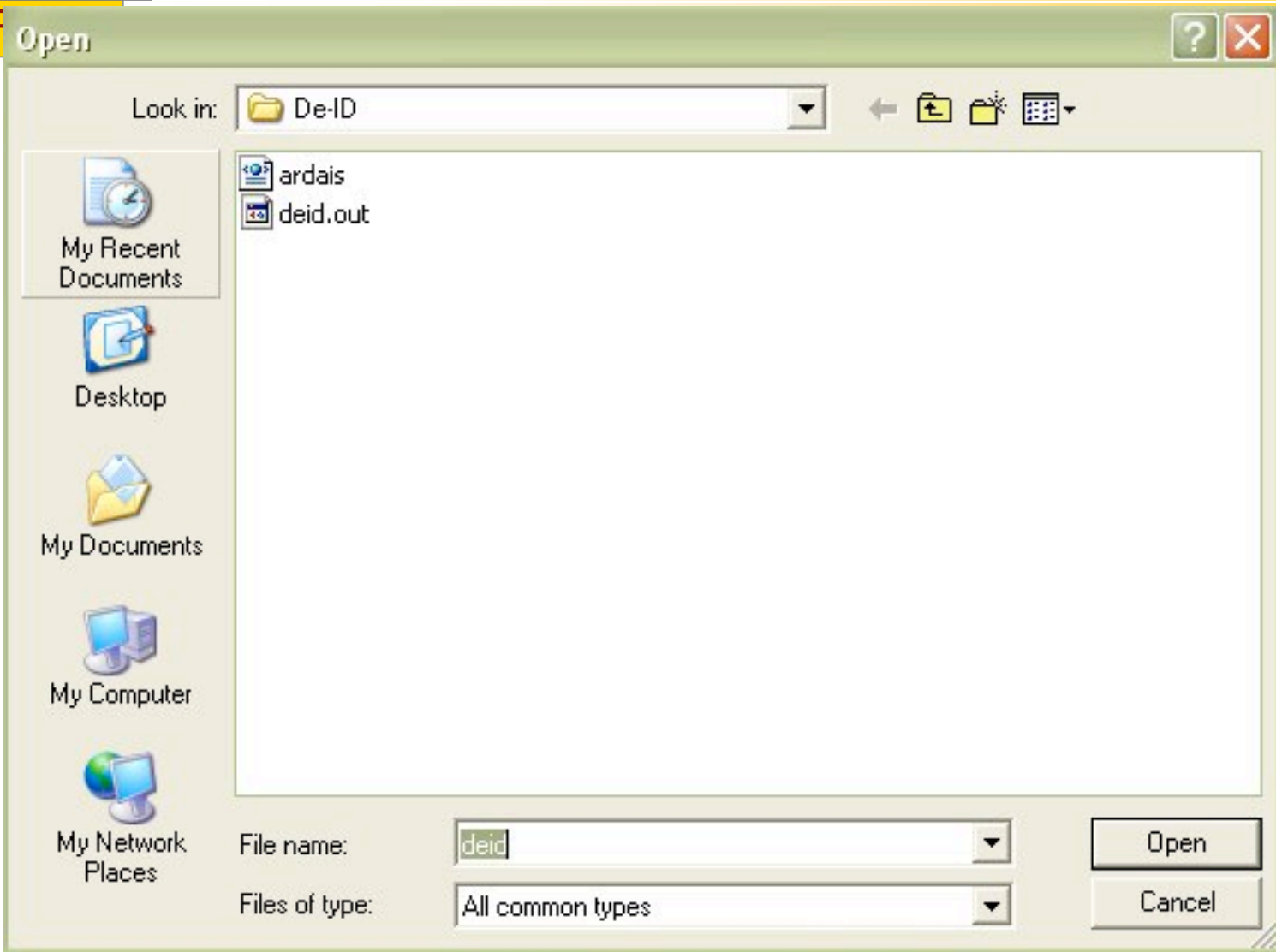
Safe

Text (multi-line header)
Text (multi-line header)
Text (tabbed header)
MARS-BSH
Tab Delim
XML

De-ID

DATA

CONFIDENTIAL



CONFIDENTIAL
Not for Publication or Distribution

De-ID DATA

CC

De-ID / Software for de-identifying narrative medical documents

File Options Help


Input file (with identifiers)
S:\De-ID\vardais.xml
File Type: XML

☒ **Output file (de-identified)**
S:\De-ID\vardais.deid

☐ **Linkage file**
S:\De-ID\vardais.deid.lin
Passphrase:

☒ **De-identification log**
S:\De-ID\vardais.deid.log

[Limited dataset compliant Tab Delim](#)

 **Begin de-identification**

Available Header Fields:

Header Field	Type
Main Patient Identifier	ID
Record Type	Report Type
Edit Date	Date

Selected Fields:

Header Field	Type
--------------	------

De-ID copyright (c) 1999-2004, University of Pittsburgh. All rights reserved. Ver 6.00 (Jan 11 2005)

CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA

CC

De-ID / Software for de-identifying narrative medical documents

File Options Help


Input file (with identifiers)
S:\De-ID\deid.out
File Type: MARS-BSH

☒ **Output file (de-identified)**
S:\De-ID\deid.deid

☐ **Linkage file**
S:\De-ID\deid.deid.lin
Passphrase:

☒ **De-identification log**
S:\De-ID\deid.deid.log

[Limited dataset compliant Text \(multi-line header\)](#)

 **Begin de-identification**

Available Header Fields:

Header Field	Type
Checksum	MARS Cks
Main Patient Identifier	ID
Patient Name	Name
Record Type	Report Typ
Parsing Queue Transactio...	ID
Principal Physician	Name
Subgroup Classifier	Other

Selected Fields:

Header Field	Type
Principal Date	


Principal Date
De-identify header item
Don't de-identify header item

De-ID copyright (c) 1999-2004, University of Pittsburgh. All rights reserved. Ver 6.00 (Jan 11 2005)

CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA
CORP

De-ID



deid.d

100%

[Show Diagnostics](#) | [Close Window](#)

De-ID
DATA
CORP

Text Output 1

S_O_H
Counters Report Type
1,pJYQ49wgYF+h SP
E_O_H
[Report de-identified (Safe-harbor compliant) by De-ID v.6.00]

PATIENT HISTORY:

The patient is a **AGE[in 20s]-year-old female with mole changing color and growing larger.

OSS, ***PATH-NUMBER[1], **DATE[May 18 2003]

PRE-OP DIAGNOSIS: Mole of back.

POST-OP DIAGNOSIS: Not given.

/EVA

FINAL DIAGNOSIS:

SKIN, BACK, BIOPSY (OSS, ***PATH-NUMBER[1], **DATE[May 18 2003])

- A. NON-ULCERATED SUPERFICIAL SPREADING MALIGNANT MELANOMA ARISING IN A BACKGROUND OF A DYSPLASTIC NEVUS.
- B. CLARKE'S LEVEL III; 1.7MM 1-2 MITOSIS PER 10 HIGH-POWER FIELD.
- C. BRISK MULTIFOCAL TUMOR INFILTRATING LYMPHOCYTES.
- D. PATHOLOGIC TNM STAGING : ~~DT~~ NM.

**INITIALS

Pathologist: **NAME[WWW **NAME[M. M. XXX], M.D.

Fellow/Chief Resident: **NAME[SSS M. RRR], M.D.

** Report Electronically Signed Out **

By Pathologist: **NAME[WWW **NAME[M. M. XXX], M.D.

**DATE[May 30 2003] 11:56

My signature is attestation that I have personally reviewed the submitted material(s) and the final diagnosis reflects that evaluation.

CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA
CONF

Text Output 2

```
OUTSIDE ACCESSION #:  
2 CONSULT SLIDES LABELED ***PATH-NUMBER[1]  
2 CONSULT BLOCKS LABELED ***PATH-NUMBER[1]  
OUTSIDE REPORT RECEIVED: Y  
CONSULT MATERIAL DESCRIPTION:  
Received for consultation from **NAME[M. ZZZ], M.D., are two (2) consult slides  
and two (2) consult blocks labeled ***PATH-NUMBER[1] from **PLACE,  
Pathology Department, **PLACE, PA, along with an accompanying pathology  
report.  
/SVM  
MICROSCOPIC:  
    SYNOPTIC DATA - PRIMARY CUTANEOUS MELANOMA  
    Family history of melanoma is unknown.  
    Family history of dysplastic nevi is unknown.  
TUMOR LOCATION:      Back  
TYPE OF PROCEDURE:    Excisional biopsy  
SIZE OF TUMOR:        Maximum surface diameter of neoplasm: 8 mm  
GROSS ULCERATION:      No  
GROSS SATELLITES:      No  
HISTOLOGIC TYPE:      Superficial spreading  
SURFACE ULCERATION:    Not present  
CLARK'S LEVEL: III  
BRESLOW'S THICKNESS:   1.7 mm  
ANGIOLYMPHATIC INVASION:  None  
PERINEURAL INVASION:   No  
EVIDENCE OF REGRESSION: No  
MICROSCOPIC SATELLITES: No  
PREEXISTING NEVUS:     Preexisting nevus present  
TUMOR INFILTRATING LYMPHOID INFILTRATE:  Marked (Brisk)  
MITOTIC RATE:          1 / 10 HPF  
VASCULARITY:           Moderately increased  
SURGICAL MARGIN INVOLVEMENT:  Deep margin is free of tumor, Lateral margin  
is free of tumor  
SENTINEL LYMPH NODE MAPPING:  No  
COMPLETION DISSECTION:  No  
T STAGE, PATHOLOGIC:     pt2a  
N STAGE, PATHOLOGIC:     DNX  
M STAGE, PATHOLOGIC:     DNX  
-----
```

FINAL
ITION

Study ID: pJYQ49wgYF+h

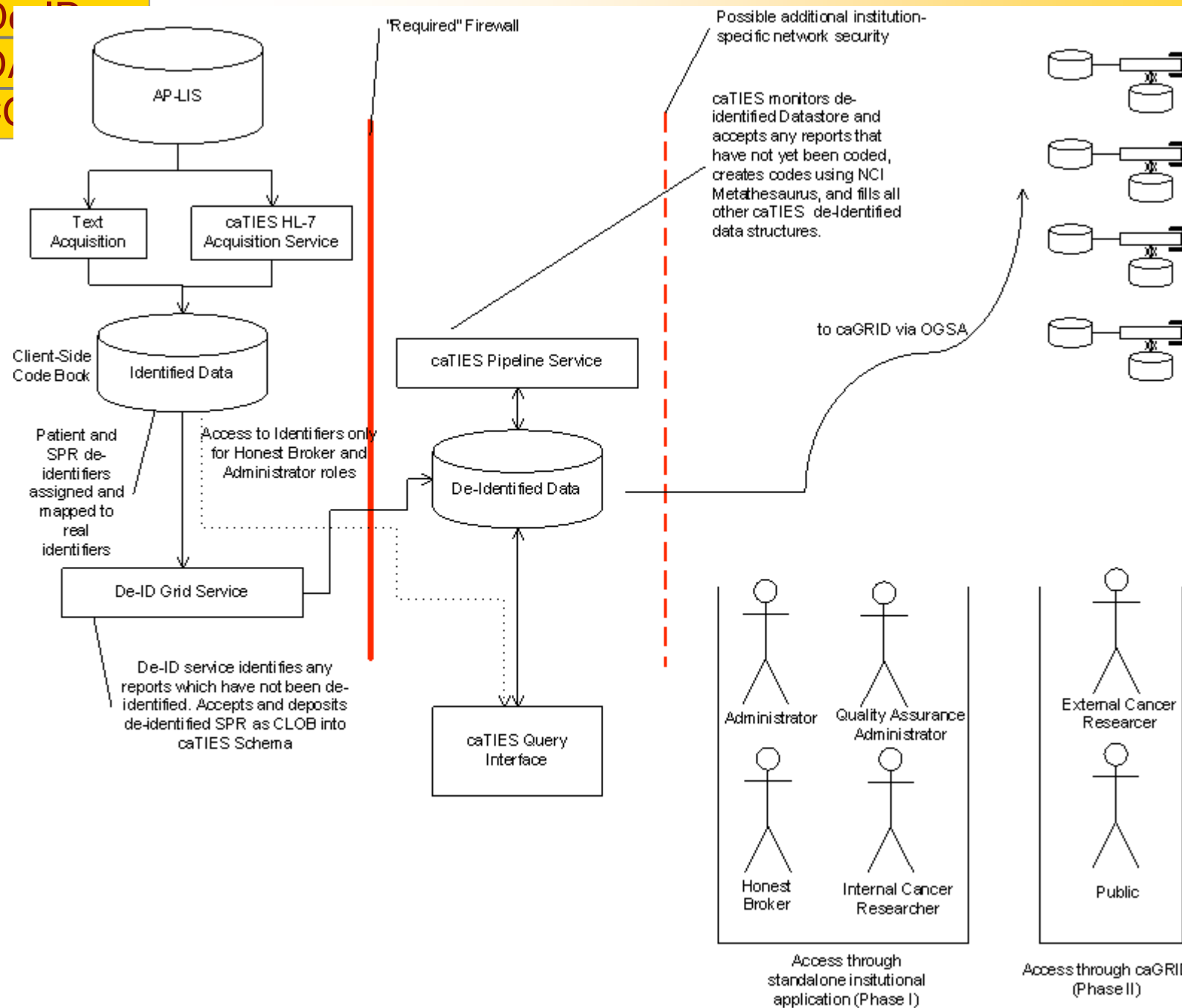
Report Type: SP

[Report de-identified (Safe-harbor compliant) by De-ID v.6.00] PATIENT HISTORY: The patient is a **AGE[in 20s]-year-old female with mole changing color and growing larger. OSS, ***PATH-NUMBER[1], **DATE[May 18 2003], PRE-OP DIAGNOSIS: Mole of back. POST-OP DIAGNOSIS: Not given. /evm

FINAL DIAGNOSIS:
SKIN, BACK, BIOPSY (OSS, ***PATH-NUMBER[1], **DATE[May 18 2003]) A. NON-ULCERATED SUPERFICIAL SPREADING MALIGNANT MELANOMA ARISING IN A BACKGROUND OF A DYSPLASTIC NEVUS. B. CLARKE'S LEVEL III; 1.7MM 1-2 MITOSIS PER 10 HIGH-POWER FIELD. C. BRISK MULTIFOCAL TUMOR INFILTRATING LYMPHOCYTES. D. PATHOLOGIC TNM STAGING : pT NM.
**INITIALS Pathologist: **NAME[WWW **NAME[M. M. XXX], M.D. Fellow/Chief Resident: **NAME[SSS M. RRR], M.D. ** Report Electronically Signed Out ** By Pathologist: **NAME[WWW **NAME[M. M. XXX], M.D. **DATE[May 30 2003] 11:56 My signature is attestation that I have personally reviewed the submitted material(s) and the final diagnosis reflects that evaluation.

OUTSIDE ACCESSION #:
2 CONSULT SLIDES LABELED ***PATH-NUMBER[1] 2 CONSULT BLOCKS LABELED ***PATH-NUMBER[1] OUTSIDE REPORT RECEIVED: Y CONSULT MATERIAL DESCRIPTION: Received for consultation from **NAME[M. ZZZ], M.D., are two (2) consult slides and two (2) consult blocks labeled ***PATH-NUMBER[1] from **PLACE, Pathology Department, **[PLACE] PA, along with an accompanying pathology report. /evm MICROSCOPIC: SYNOPTIC DATA - PRIMARY CUTANEOUS MELANOMA Family history of melanoma is unknown. Family history of dysplastic nevi is unknown. TUMOR LOCATION: Back TYPE OF PROCEDURE: Excisional biopsy SIZE OF TUMOR: Maximum surface diameter of neoplasm: 8 mm GROSS ULCERATION: No GROSS SATELLITES: No HISTOLOGIC TYPE: Superficial spreading SURFACE ULCERATION: Not present CLARK'S LEVEL: III BRESLOW'S THICKNESS: 1.7 mm ANGIOLYMPHATIC INVASION: None PERINEURAL INVASION: No EVIDENCE OF REGRESSION: No MICROSCOPIC SATELLITES: No PREEEXISTING NEVUS: Preexisting nevus present TUMOR INFILTRATING LYMPHOID INFILTRATE: Marked (Brisk) MITOTIC RATE: 1 / 10 HPF VASCULARITY: Moderately increased SURGICAL MARGIN INVOLVEMENT: Deep margin is free of tumor, Lateral margin is free of tumor SENTINEL LYMPH NODE MAPPING: No COMPLETION DISSECTION: No T STAGE, PATHOLOGIC: pT2a N STAGE, PATHOLOGIC: pNX M STAGE, PATHOLOGIC: PMX -----

De-ID
Data
CO



De-ID

DATA

CORP

Summary

- Network-level operability
- Automated batch feed, output management
- High volume
- Rapid, reliable
- Leaves original file intact
- Outputs ready for parsing, tagging



De-ID
DATA
CORP



Optional Re-Identification

De-ID

DATA

CORP

Re-identification

- Option to create linkage file
- Linkage file independent of de-identified data
- Linkage files are encrypted and password protected

De-ID DATA

CC

De-ID / Software for de-identifying narrative medical documents

File Options Help


Input file (with identifiers)
S:\De-ID\deid.out
File Type: MARS-BSH

☒ **Output file (de-identified)**
S:\De-ID\deid.deid

☒ **Linkage file**
S:\De-ID\deid.deid.lin
Passphrase: abc123

☒ **De-identification log**
S:\De-ID\deid.deid.log

[Limited dataset compliant Tab Delim](#)

 **Begin de-identification**

Available Header Fields:

Header Field	Type
Checksum	MARS Cks
Main Patient Identifier	ID
Patient Name	Name
Principal Date	Date
Parsing Queue Transactio...	ID
Principal Physician	Name
Subgroup Classifier	Other

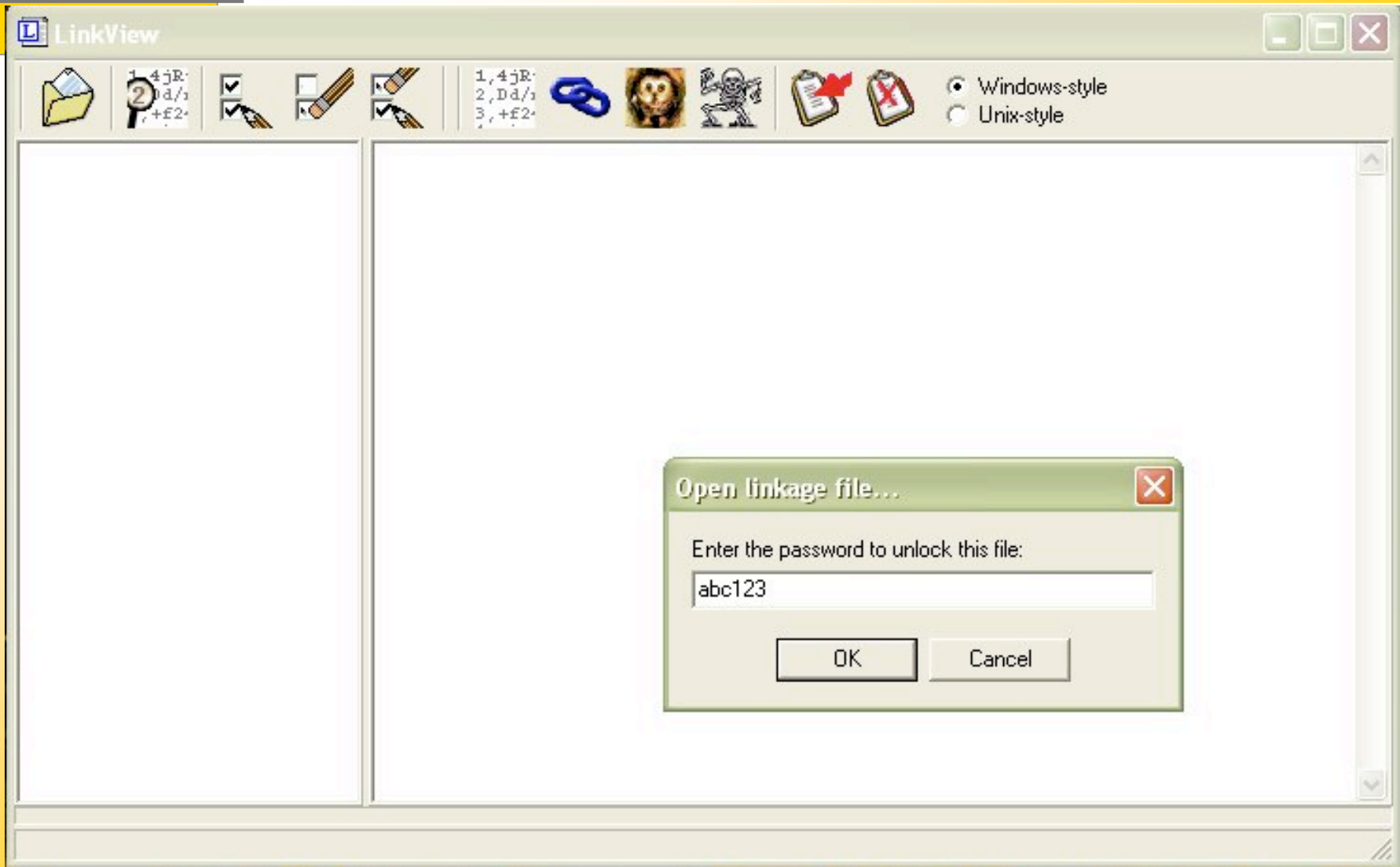
Selected Fields:

Header Field	Type
Record Type	Report Type

De-ID copyright (c) 1999-2004, University of Pittsburgh. All rights reserved. Ver 6.00 (Jan 11 2005)

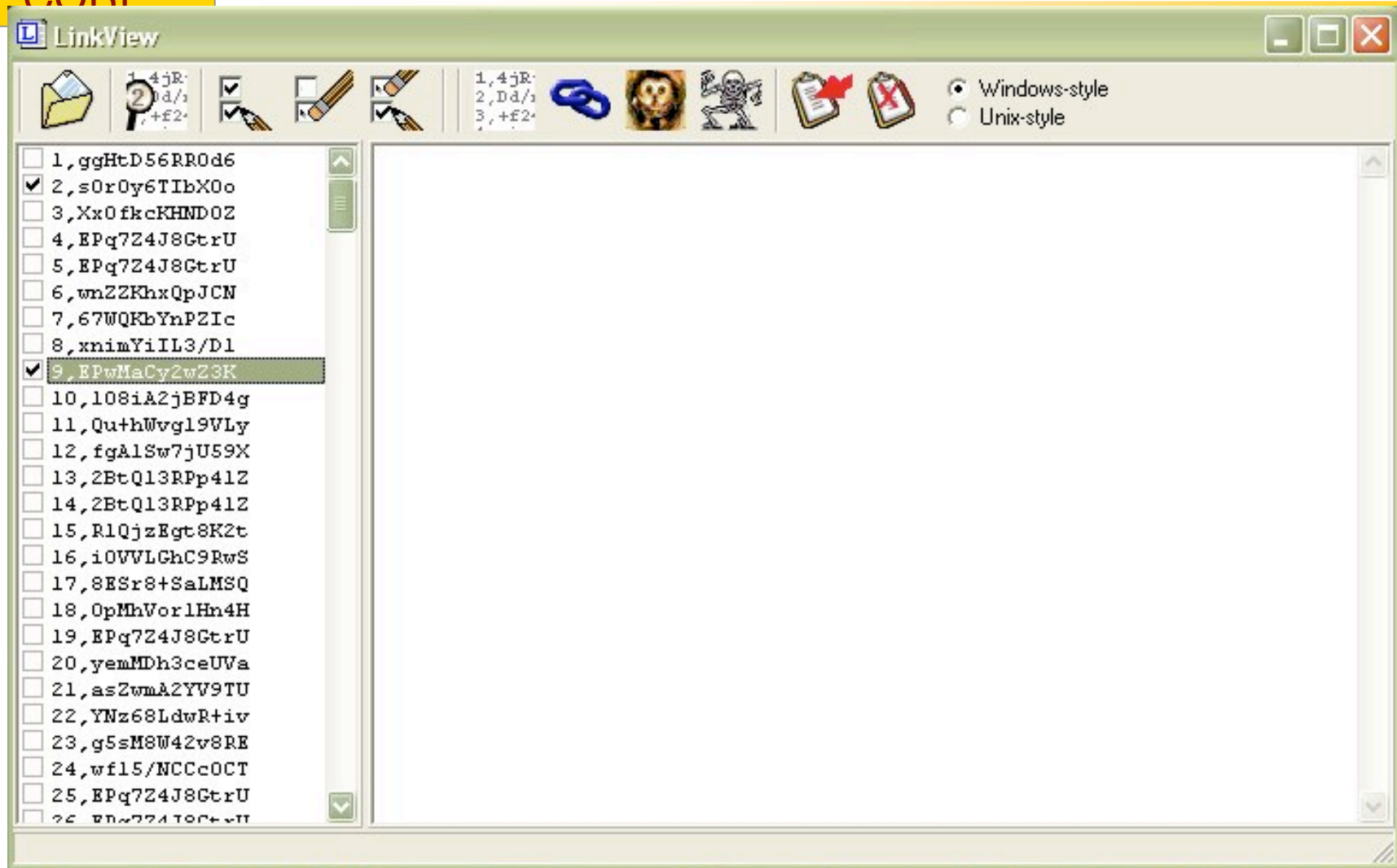
CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA



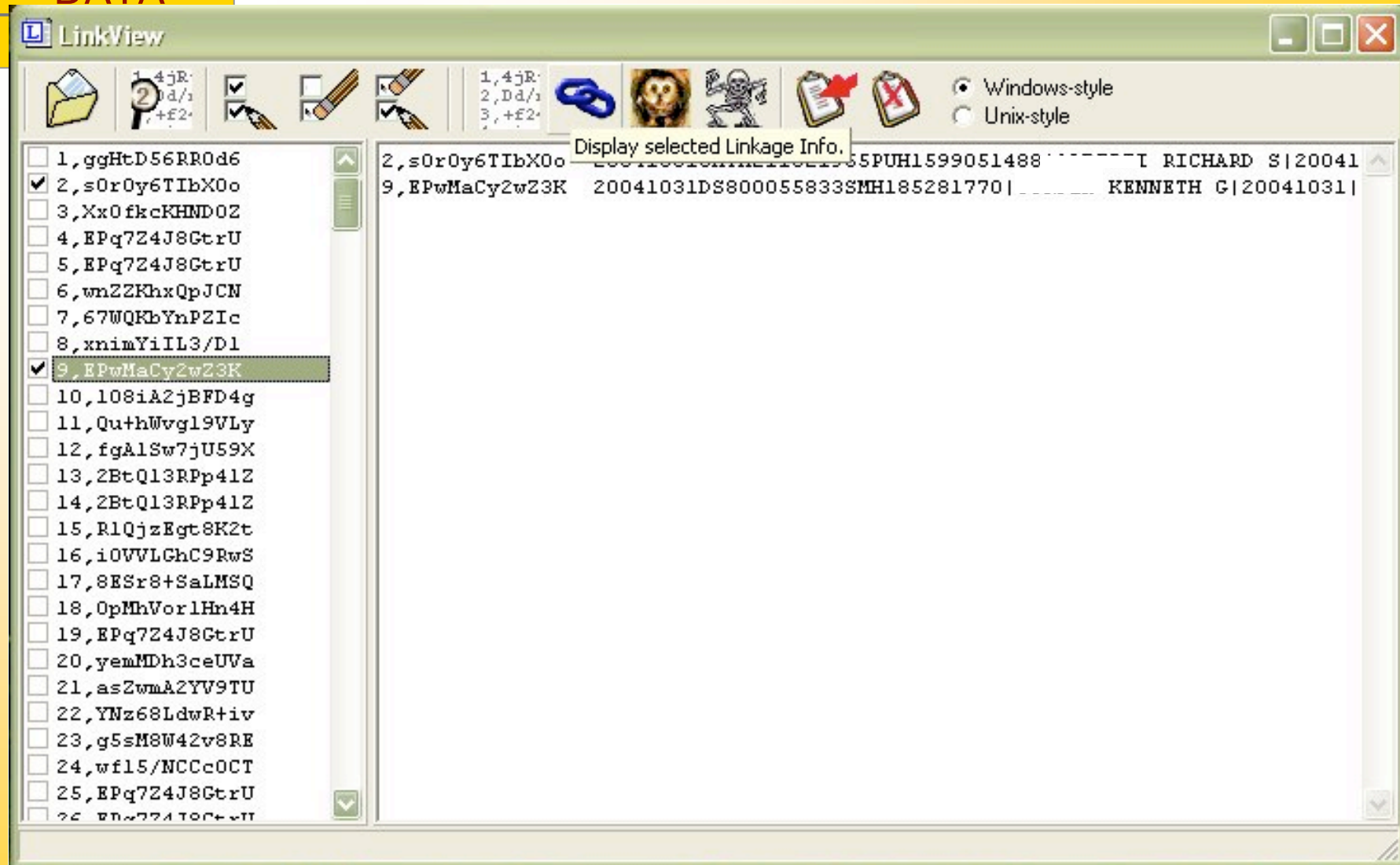
CONFIDENTIAL
Not for Publication or Distribution

De-ID
DATA
CORP



CONFIDENTIAL
Not for Publication or Distribution

De-ID DATA



CONFIDENTIAL
Not for Publication or Distribution